DOCUMENT RESUME

ED 132 201 TM 005 938

AUTHOR Hyman, Irwin A.; And Others
TITLE Developing Criterion Referenced Assessment for Head Start: Theoretical and Practical Considerations.
NOTE 17p.; Expanded version of a paper presented at the Annual Meeting of the American Psychological Association (Washington, D.C., September 3, 1976)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS Bilingualism; *Compensatory Education Programs; *Criterion Referenced Tests; *Disadvantaged Youth; Early Childhood Education; Intelligence Tests; Minority Group Children; Nature Nurture Controversy; Norm Referenced Tests; Norms; Predictive Validity; *Screening Tests; Standardized Tests; Test Bias; *Testing Problems; Test Reliability
IDENTIFIERS *Project Head Start

ABSTRACT

Testing with minority and disadvantaged populations in America has resulted in widespread abuse. Theorists have historically viewed tests as instruments for upward mobility in a system in which doors are often opened by parental wealth and status. Predictive validity of tests has not properly accounted for problems in correlational techniques which accrue when low socioeconomic minority groups who tend to live in homogeneous clusters are considered within the general population. One solution to the lack of norm referenced groups which reflect the homogeneity of minority groups is the use of criterion referenced tests which are useful for both diagnosis and teaching. A criterion referenced approach has been developed over a five year period. The results have been used for pre and post measures for assessing efficacy, for individual instruction, for screening children for further assessment and as a general guide to help teachers understand individual needs. The test was developed by the Head Start staff and is periodically updated. Factor analysis suggests a specific language factor and a general readiness factor. Internal consistency has run consistently over .88 for administration by college sophomores. Problems of design and implementation are discussed. (Author/RC)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED132201

TM005 938

DEVELOPING CRITERION REFERENCED ASSESSMENT FOR HEAD START: THEORETICAL AND PRACTICAL CONSIDERATIONS[1]

Irwin A. Hyman
Temple University

Elinor Chafey
Trenton State College

Kathleen Smith
Temple University

It is difficult for some to understand the current attack on testing. Objective scientific assessment procedures have, for many, provided a means of advancement in our "technocracy." For many, test results have permitted upward mobility across socioeconomic class lines and minimized the affects of racial and ethnic prejudices on personal advancement.

"Proponents of testing, from Thomas Jefferson onward, have wanted to open doors for the talented poor, in a system in which doors often are opened by parental wealth and status" (Cronbach, 1975). However, there is currently widespread dissatisfaction with the use of tests and the psychological community cannot ignore either the substantive agreements or the growing public prejudice against their use.

The furor which has developed around the issue of testing minority children has resulted in a wide range of actions including a proposed moritorium on I.Q. tests (Williams, 1974). It is the belief of many, however; that such action is inappropriate (APA, 1966; Fishman, et al, 1964; Milgram, 1974; Newland, 1974). Perhaps it is not the tests themselves which should

[1] This paper is an expanded version of the paper presented at the Annual Meeting of the American Psychological Association, Washington, D.C., September 3, 1976.

be faulted, but how the results are interpreted and used (Darlington, 1971; Wickoff, 1974). However, despite rational arguments about the misuse of tests and suggested changes (Cleary, Humphreys, Kendrick, and Wesman, 1975) the harmful use of test scores with disadvantaged students has generated increased resistence to the use of any standardized test with any group (Adams, 1976; Davis, 1976).

The development of the testing industry in America has in many ways paralleled the development of other technologies in our society. In brief, scientists discover basic theoretical constructs which are then tested and often later applied to real life problems. The history of America has been characterized by the often too rapid development of technologies which eventually create more problems than they solve. The testing movement now faces the same type of consumer resistence, legislative concern and internal professional debates which have characterized the development of nuclear energy, the use of preservatives in foods, etc. Unfortunately, from a professional point of view, much of the criticism of the uses of tests is valid. There can be little question that improper use of tests has caused great harm in industrial, educational and instututional settings. In many cases, the tests are used inappropriately and contrary to the original intent of the test developers. As a result, psychologists are now subject to a vast onslaught on the use of tests in various settings where they have been rigidly institutionalized. There is little doubt that great mischief has been done to minority group children in lower socio-economic classes when the test results have been used only for placement and classification rather than for diagnosis and treatment. Unfortunately, this movement of resistance has grown to such proportions that it is difficult for

legislatures, judicial groups and the public in general to separate legitimate from false issues. As a result, there is a tendency to "throw the baby out with the water."

A major theoretical problem in testing disadvantaged groups is created by current standardization procedures. The current practice is to include in the standardization group a percentage of each minority or disadvantaged population that is representative of that group in the total population. Two possible solutions include the development of separate norms and the use of criterion referenced tests with disadvantaged groups. A theoretical argument for the use of separate normative and special modifications for tests with disadvantaged is presented. This is followed by a description of the use of a criterion referenced approach with one particular group of children.

The problem addressed herein is that testing is necessary and can be appropriate in assessing the strengths and weaknesses of a child so that a realistic educational program can be developed to meet the needs of the individual. Standardized intelligence tests have proved over time to be reliable and valid predictions of achievements in our middle class schools. Both the Stanford Binet and the Wechsler Intelligence Scale have recently been restandardized to reflect a more representative sample of the United States population. Implicit in this restandardization was the attempt to incorporate representative and proportionate numbers of minority children. However, procedures such as these may be faulty, especially in relation to Blacks, since they have "presumed a fairly homogeneous distribution of population within the various social classes and thus have ignored the heavy concentration and over-representation of Blacks in the lower socioeconomic levels and thus, norms established did not provide for an adequate representation of the black population (Samuda, 1975, p.6)." In testing minority children, it appears that the restandardizations have not adequately counterbalanced other factors

which influence test performance. Higher numbers of minority group children continue to fall within the lower ranges of the intelligence scales (Samuda, 1975). Some characteristics of children from lower socioeconomic levels which may be expected to affect test performance, have been identified by Fishman, et al, (1974) and are as follows, "In contrast to the middle class child, the lower class child will tend to be less verbal, more fearful of strangers, less self-confident, less motivated toward scholastic and academic achievement, less competative in the intellectual realm, more irritable, less conforming to middle class norms of behavior and conduct, more apt to be bilingual, less exposed to intellectually stimulating materials in the home, less varied in recreational outlets, less knowledgable about the world outside his immediate neighborhood and more likely to attend inferior schools." It is not enough to be sophisticated in testing, one must also take into account the influences of the child's socioeconomic status and cultural experiences upon the testing process and outcome.

It is at this point that the issue of hereditary/environmental influences on intelligence should be briefly addressed. It will be done in a very simplistic and general way which is not to ignore the complexity of issues within each theoretical approach. Jensen (1969), as a representative of the hereditarian theoretical view, has maintained that intelligence is primarily genetically determined, that in fact, for whites, intelligence is a highly hereditable trait and thus environmental factors have very little influence upon the variance of intelligence in the general population and specifically among races. Jensen's position continues to be a highly debated one, and it is not the function of this paper to support or dispute it. However, it is the position of this paper that hereditary and environment <u>both</u> must be considered

in assessing the intellectual potential of an individual, thus the environmental theoretical approach is accepted. Assumed herein is the belief that there exist great differences between the environmental conditions, experiences and knowledge according to socioeconomic class (Deutsch & Deutsch, 1976; Drake, 1966; Samuda, 1975) and that environmental factors are critical in understanding the variability in the measured intellectual ability on minority children (Baratz & Baratz, 1970; Gussow, 1974; Katz, 1968; Whiteman & Deutsch, 1967).

Numerous alternatives have been explored in an effort to develop more equitable techniques of assessing the minority child's intellectual ability and achievement potential. Samuda (1975) devotes an entire chapter of his book in discussion of culture free/culture fair tests, measures of the environment, criterion referenced testing, etc. However, it is suggested here that other alternatives, utilizing the highly respected, established intelligence tests are possible. What is positioned here is similar to the position of Fishman et al (1964). "Standardized tests, when used with disadvantaged minority groups, (1) may not provide reliable differentiation in the range of minority group scores, (2) their predictive validity for minority groups may be quite different from that for the standardization and validation groups, and (3) the validity of their interpretations is strongly dependent upon the adequate understanding of the social and cultural background of the group in question (p.130)."

It is a fact that in the United States, lower socioeconomic minority groups, through segregation tend to live in homogeneous clusters, and by virtue of their minority group status form a homogeneous socioeconomic subgroup within the greater population. When "the tester attempts to make differentiations within the group which is more homogeneous than the reference or norm group for which reliability is reported, the actual effectiveness

of the test will tend to be lower than the reported reliability coefficient appears to promise" (Fishman et al, 1964, p. 131). The consistently observed smaller spread of scores primarily within the lower range of the standardized intelligence tests draw attention to the concern that the intelligence tests are not truly differentiating between lower and/or differences. The narrower the range of scores, the lower the coefficient of reliability (Wesman, 1952). Predictive validity is undermined through the variability of cultural experiences as well as other test-related factors (Fishman, et al, 1964).

It appears that what is needed at the present time are new norm referenced groups which reflect the homogenity of minority group children upon which the intelligence tests can be restandardized. Non-English speaking minority children constitute a specific concern due to the fact that language translation techniques do not necessarily assure that the tests are language adaptive or that they provide for the additional handicap of lack of mastery of language due to bilingual confusion between home and school (Lennon, 1969; Samuda, 1975). It has been found that bilingualism has adversely affected test performance and that children who are in the process of learning English score lower than those who have mastered it (Anastasi & Cordova, 1953). New norms could be used concurrently with the established norms, but would serve to be a more discriminating estimate of intellectual ability which would be less susceptible to cultural variability.

Besides the possible use of new norms, there is a need to continue to work towards the development of more culture fair tests and <u>other</u> valid assessment procedures.

One approach, suggested by Cleary, Hymphreys, Kendrick and Wesserman (1975) is the use of criterion referenced assessment. A theoretical rationale for this approach is that it avoids the need for large samples needed for adequate standardization, it eliminates the need to compare children with each other, and it is geared to teaching without labeling children. It avoids many of the

pitfalls described previously in regard to I.Q. tests or standardized achievement tests yet it may be very useful to help in the learning process of disadvantaged children without providing unwarranted stigmatization or self-fulfilling prophecies by teachers.

About six years ago, the senior author was employed as a consultant in Trenton, New Jersey to both the Head Start and Follow Through programs. In order to coordinate the efforts of both programs, which were independently funded, he attempted to identify early, those children in Head Start who needed special attention in terms of academic, social or emotional growth. At that time, the Follow-Through model, which is based on the behavior analysis approach of Don Bushell, was utilizing a screening test for placement of children at various levels in programmed curriculum materials. The screening instrument was used to assess the criteria expected of children who would enter first grade with the necessary readiness skills for learning to read. Despite the fact that there are many disadvantaged children in the Trenton schools, it was found that the regular curriculum utilized in the typical middle-class school system formed the basis of expectations for children in Trenton. Because it appeared difficult, if not impossible to change the system's unrealistic expectations, it was felt that Head Start had a role in preparing children to meet those realities which they must face when they enter the public school. The original screening instrument was modified extensively in order to realistically reflect the program potential of Head Start. This was done through an examination of developmental literature and consultation with the Head Start staff. After the initial form of the test was developed, the question of administration to over 200 children was considered. During the initial year, the only staff available to do this was either the teachers or the central office staff which was made up of social service and medical aides. It was decided to use the aides since there are many complications from this writer's experience in utilizing teachers to obtain objective

test information. The aides were trained during five one-hour sessions. Children were tested within a four week period near the end of the school year. Children were familiar with the aides who had been in and out of the classroom for various reasons. Each child was assessed individually in either a separate room or a quiet corner of the Head Start area. The technique of using the aides had several advantages since the testers were known to the children and they were representative of the minority group. Since they were always available, they could later test children who were frequently absent. The obvious disadvantage of using the aides was the time involved which kept them from their regular duties. Also, situations varied so that sometimes aides tested children they knew quite well and liked, and as a result they were biased in favor of those children. The opposite was also true. This program took approximately 2 to 3 full weeks of their time and although the initial effort was successful, it was felt that it was economically unfeasible to continue this approach.

The following year, we attempted to enlist the aid of volunteer groups to be trained to administer the test. This became extremely difficult since we wanted the test administered during a relatively short period of time so that the results would be useful in curriculum planning. After exhausting a variety of possibilities for test administration, a fortuitous contact was made with a local teacher's college. The utilization of the college is mutually beneficial. The students, who are early childhood majors, are presented with the opportunity of administering a developmental assessment instrument with young children. This provides an excellent opportunity for the student's classroom learning to be understood in terms of practical applications. For many of the middle-class students, this is their first experience with children in economically disadvantaged populations. As a result of this mutual interchange, many students realize that they are functioning under concepts that are more myth than reality in relation to

these children. The obvious advantage for the Head Start program is the provision each fall and spring of a group of highly motivated people, interested in young children who may be easily trained to administer the assessment device. Another distinct advantage is that with a large number of students, the testing can be completed within a one-week period. This allows for administrative ease and permits comparisons of data over a number of years. Initially, students are trained by the psychologist or his interns in school psychology well in advance of the testing at the Head Start Centers. This training consists of introduction to the test and a demonstration of its administration. The students' college instructor further discusses the relationship between theoretical variables and the actual test protocol. Students are then provided an opportunity to administer the test to children in the experimental nursery school which is associated with the teachers college.

The test was developed for relatively easy administration within a 20 to 30 minute period per child. Training is covered within a 3-week period. Preparation for testing is made far in advance with the Head Start Centers, since there are many scheduled events in the weekly calander of Head Start. Students are assigned to centers on a basis of need and in the two centers where many Spanish speaking children are housed, Spanish speaking students are assigned to use the Spanish protocol. Students are assigned to centers for specific blocks of time which are arranged between the teachers and the coordinator of the testing. They are requested to spend some time in the centers before beginning testing in order to familiarize the children with them and in order to get some idea of the developmental level of the children whom they are going to be testing. These are then assigned to students as they arrive. The protocols are utilized for marking comments and scoring the test. In addition, mark scanning sheets are used to record the scores. Testers take the test protocols and the mark scanning sheets with them and they are double checked. They are handed in as

part of the class assignment. Within an approximately 3 week period, the protocols with the actual responses are returned to the teachers and the mark scanning sheets are utilized for computer analysis and individual scoring of the tests.

The tests are usually administered in October and May so that teachers have an opportunity to evaluate each child's score and test performance in the various areas of the test at the beginning of the year and towards the end. When test results do not agree with the teacher's assessment of the child, teachers are encouraged to retest the child, since they have additional test protocol and booklets in the centers. Teachers use the individual test performance of children as a guideline in helping towards readiness training. The teacher involvement in the planning and utilization of the test scores demystifies the concept of testing and helps the teachers to realize that items on the test only represent a sample of the types of criteria which are important for children to attain in order to prepare for public school. When children are particularly weak or strong in specific areas, this is noted. The protocols and the follow-up teaching provides an excellent format which is utilized by teachers in their periodic home conferences with parents. This is especially helpful where the teachers are able to outline specific kinds of activities for parents to work on with children when this is indicated.

This paper has presented a rather detailed account of the use of a criterion referenced test. The test was developed in cooperation with the teachers who would use the results. It is based on criteria most of which are obtainable by the majority of children in the program. The test is easily and inexpensively administered and the results are used for diagnosis and educational planning. The test is a screening instrument and the teachers have the items easily available for readministration.

The limitations of the test are recognized by both the developer and the users. The test results are not in any way used to classify or categorize children, since it is recognized that this is a screening instrument. Since the items are based on many educational tasks required of children between the ages of 3½ to 5½, this type of test also provides a developmental screening instrument for children at the preschool level. There has been a great deal of controversy over the concept of developmental screening. The E.D.S.P.T. program has mandated developmental assessment for children who are enrolled in programs like Head Start. However, the generally recommended Denver Developmental Test has many limitations, including the major one that it is based on national norms. The problem of national standardization was discussed at the beginning of this paper in reference to I.Q. tests. Further, it is a rather gross instrument and does not lend it self to utilization for cognitive assessment and remediation. Therefore, the present criterion referenced approach was conceptualized as also offering a norm referenced approach, if utilized adequately. In Trenton, the test has been used to generate local norms. During initial screening, the statistics generally are normally distributed, and showing a slight negative skewness. Over several years of use, the internal consistency measure using the Kuder-Richardson 20 has been over .900. Although the test is divided into 11 areas organized conceptually, initial factor analysis after rotation suggested two main factors. A common readiness factor and a separate specific expressive language factor as measured by children's ability to identify pictures. The scoring procedure on the OPSCAN I.B.M. sheets can be easily adopted to a format used by most colleges for scoring objective tests. It is relatively simple to use a canned program at any college that provides a test service bureau. In this particular situation, the OPSCAN sheets are easily and quickly processed to generate complete statistical data plus individual name, scores, etc.

The percentile scores are used to identify children at the upper and lower ends of the continuum. Specific children are then followed up by further psychological or educational and medical evaluation as indicated. For instance, the children in the 95th percentile and above, are identified through this procedure. Teachers are requested to provide further information on these children. Follow-up consists of classroom observation and perhaps further psychological testing to assess intellectual ability. This procedure further enables the psychologist and teacher to plan individual programs in Head Start for these particular children. At the end of the year, those children identified as being far advanced are identified for the kindergarten so that they do not spend another year doing the same things that they have already accomplished.

Children at the lower end of the continuum have usually been identified by their teachers through observation as being relatively slow or having other handicaps. These children, of course, receive further assessment and are then identified as having special needs if the case warrants this designation. Although there have been many complaints about the early classification of children, it must be recognized that in order for children to receive special help through most current legislation, they must be identified and classified. Specific differential diagnosis excepting in clear medical cases is usually avoided. The focus is on treatment procedures in the classroom and home utilizing the criterion referenced test results as a guideline.

This paper has presented a method for using both norm referenced and criterion referenced approaches with the same test, at the local level. Involvement by all staff, parents and the psychologist assures that adequate input is provided so that no child is unjustly, unfairly or ill-advisedly designated as having a special problem. Despite the efforts described, it would be untruthful to say that all has been rosy. There have been some attacks on the testing program. Initially, these came from a few parents who were knowledgeable about the bad

effects of testing Black children living in disadvantaged areas. However, regular meetings with the parents before each testing period has resulted in no significant problems from that sector. Parents want their children to be competent and when the test is explained, they are glad to see that there will be some way of assessing what the children need to learn and what they learn during the year. However, in this particular program, one staff member and a consultant made a concerted attack on the testing program based on the assumption that all testing, except that by teachers, tends to be harmful to children (Holmes and Nolte, 1976). Because the consultant became involved in what was essentially internal conflicting philosophies, the program staff was unable to resolve the problem before it came to the attention of the community representative from the Regional Office. Despite the fact that in essence, with minor problems, almost all staff and parents were satisfied, a major effort was extended by professionals in maintaining the program before the final resolution. This type of resistence must be anticipated in any assessment program with minority groups. However, despite the energy and often bitter hours spent in battle, the program has survived and continued to be modified. The community at large, especially the parents, have found the approach most beneficial in helping their children prepare for kindergarten, which is what Head Start is all about.

REFERENCES

Adams, V. The making of tests: a very inexact science. The New York Times, May 16, 1976, pp. E11.

American Psychological Association. Standards for Educational and Psychological Tests. New York: American Psychological Association, Inc., 1966.

Anastasi, A. & Cordova, F.A. Some effects of bilingualism upon the intelligence test performance of Puerto Rican children in New York City. Journal of Educational Psychology, 44, 1953, 1-19.

Baratz, S.S. & Baratz, J.C. Early childhood intervention: the social science base of institutional racism. Harvard Educational Review, 40, 1970, 29-50.

Bennett, G.K. Response to Robert Williams. The Counseling Psychologist, 2, 1970, 88-89.

Cleary, T.; Humphreys, L.; Kendrick, S.; and Wesman, A. Educational uses of tests with disadvantaged students. American Psychologist, 30, 1975, 15-41.

Cronbach, L.J. Five decades of public controversy over mental testing. American Psychologist, 30, 1975, 1-14.

Davis, D. Testing: multiple choices for citizens. Citizen Action in Education, 3, 1976, (complete issue devoted to testing).

Darlington, N.B. Another look at cultural fairness. Journal Educational Measurement, 8, 1971, 71-82.

Deutsch, M. & Deutsch, C.P. Brief reflection on the theory of early childhood enrichment programs. In M. Deutsch et al (Eds.) The Disadvantaged Child. New York: Basic Books, 1967, 195-307.

Drake, St. C. The social and economic status of the Negro in the United States. In T. Parsons and K.B. Clark (Eds.) The Negro American. Boston: Beacon Press, 1966, 3-46.

Fishman, J.A. etal. Guideline for testing minority group children. Journal of Social Issues Supplement, 20, 1964, 129-145.

Green, R.L. Tips on educational testing: What teachers and parents should know. Phi Delta Kappan, 1975, October, 89-93.

Gussow, J.D. Nutrition and mental development. Urban Disadvantaged Series, 36, 1974, 1-41

Holmes, M. & Nolte, D. A position paper on criterion referenced assessment. Unpublished paper, Trenton Head Start, February 19, 1976.

Jensen, A.R. How much can we boost IQ and scholastic achievement? Harvard Educational Review, 39, 1969, 1-123.

Karmel, L.J. Measurement and Evaluation in the Schools. New York: Macmillan Company, 1970.

Katz, Academic motivation and equal education opportunity. Harvard Educational Review, 38, 1968, 56-65.

Kimble, G.A. & Garmezy, N. Principles of General Psychology. 3rd ed. New York: The Renald Press, 1968.

Lennon, R.T. Statement of the test department, Harcourt Brace Javanovich, Inc. Regarding the Use of Standardized Instruments with Urban and Minority-Group Pupils. In a symposium in Testing as a Social Problem: Issues and Responsibilities. American Psychological Association, Washington, D.C., August 31, 1969.

Lennon, R.T. Testing: The question of bias. In T.J. Fitzgiban, (Ed.) Evaluation in the inner city. New York: Harcourt, Bruce and Wald, 1970.

Mehrens, W.A. & Lebmann, I.J. Standardized tests in Education. New York: Holt, Rinehart and Winston, 1969.

Mercer, J.R. "Institutional anglocentrism: Labeling mental retardates in the public schools." In P. Orleans and W. Russell (Eds.) Race changes and urban society, Vol. V. Los Angeles: Sage Publications, 1971.

Milgram, N.A. Danger: Chauvenism, Scapegoatism and Euphanism. In J.G. Williams and S. Gerdon (Eds.) Clinical Child Psychology. New York: Behavioral Publications, 1974.

Newland, T.E. Testing Minority Group Children. In G.T. Williams and S. Gerdon (Eds.) Clinical Child Psychology. New York: Behavioral Publications, 1974.

Patton, M.Q. What test scores really mean. Citizen Action in Education, 3, (3), 1975, 7.

Samuda, R.J. Psychological Testing of American Minorities. New York: Mead, Dudd and Company, 1975.

Whiteman, M. & Deutsch, M. Social disadvantage as related to intellectual and language development. In M. Deutsch and Associates (Eds.) The Disadvantaged Child. New York: Buner Books, 1967.

Wickoff, R.L. Danger: Attacks of testing unfair. In G.J. Williams and S. Gerdon (Eds.) Clinical Child Psychology. New York: Behavioral Publications, 1974.

Williams, R.L. Abuses and measures in testing Black children. The Counseling Psychologist, 2, 1971, 62-73.

Williams, R.L. Testing and dehumanizing Black children. In G.J. Williams and S. Gerdon (Eds.) Clinical Child Psychology. New York: Behavioral Publications, 1974, 312-315.

Williams, R.L. From dehumanization to Black intellectual genecide: In G.J. Williams and S. Gerdon (Eds.) Clinical Child Psychology. 1974, 318-319.